PLASMID DNA FROM YERSINIA PESTIS

CROSS-REFERENCE TO RELATED APPLICATIONS

Not applicable.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with United States government support awarded by _____.

BACKGROUND OF THE INVENTION

Over the centuries, the bubonic plague (also known as the Black Death) has claimed the lives of millions of people. The disease is characterized by chills, fever, vomiting, diarrhea, painful swollen lymph nodes (buboes), blackening of the skin caused by ruptured blood vessels, and a very high mortality rate (up to 75% if left untreated). Treatment with antibiotics in the early stages of the infection is generally effective.

Bubonic plague is caused by the bacterium *Yersinia pestis,* which is transmitted to humans from rats or other rodents by fleas that feed on infected rodents and then bite humans. Reservoirs of the bacteria persist today, and attempts to eliminate wild rodent plague have proven ineffective. Occasional outbreaks of the deadly disease continue to occur, particularly in small towns, villages, and rural areas in developing countries.

While bacteria carry genetic material in their chromosomes, bacteria also often carry genetic material in loops of DNA called plasmids. Bacterial plasmids are nonessential, extrachromosomal genetic elements capable of autonomous replication. The genetic material in plasmids

often encodes functions required for maintenance of the plasmid in its bacterial host and sometimes encodes optional functions that promote survival of the bacterial host under certain environmental conditions. Pathogenicity determinants are commonly plasmid-encoded, and fall within the category of optional plasmid-encoded functions.

Yersinia pestis is a facultative intracellular parasite which harbors at least three different plasmids, designated pCD1, pPCP1, and pMT1, which are necessary for full virulence of the organism. One of the plasmids, designated pCD1, is also found in the enteropathogenic species *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* (Ferber, *et al.* Infect. Immun. 31:839-841, 1981; Portnoy, *et al.* Curr. Topics Microbiol. Immunol. 118:29-51, 1985), whereas pMT1 and pPCP1 are unique to *Y. pestis* (Brubaker. Clinical Microbiol Rev. 4:309-324, 1991). Plasmids pMT1 and pPCP1 are thought to promote deep tissue penetration by *Y. pestis* and to contribute to the acute infection associated with this species. The *Y. pestis* genome shares much homology with that of *Y. pseudotuberculosis* (Bercovier, *et al.* Curr. Microbiol. 4:225-229, 1980; Moore, *et al.* Inter. J. Sys. Bacteriol. 25:336-339, 1975), yet the infection caused by the latter organism is usually mild and self limiting (Butler, Plague and other yersinia infections, p. 111-159. *In* W.B. Greenbugh III and T.C. Merigan (eds.), Current topics in infetctious diesease, Plenum Press, New York, NY, 1983).

An understanding of the differences in the pathogenesis of *Y. pestis* and *Y. pseudotuberculosis* may be afforded by comparing polynucleotide sequences or genes found on pMT1 or pPCP1 plasmids, and which are unique to *Y. pestis*. It has been found that *Y. pestis* strains lacking the pCD1 plasmid are completely avirulent. Therefore, determination of the complete pCD1 sequence may provide

important information about the role of the plasmid in virulence in various pathogenic yersiniae.

The 9.5 kb plasmid pPCP1 encodes a bacteriocin termed pesticin, a pesticin immunity protein and a plasminogen activator activity. Loss of this plasmid increases the LD_{50} of the organism by a factor of one hundred thousand, as measured by subcutaneous injection in the mouse model. (Sodeinde, *et al.* Science 258:1004-1007, 1992).

The second plasmid unique to *Yersinia pestis*, designated pMT1, is a 100-kb plasimd that encodes the capsular protein Fraction 1 and the murine toxin (Protsenko, *et al.* Genetika 19:1081-1090, 1983). The genes for the capsular proteins have been cloned and sequenced using *Y. pestis* strain EV76 (Galyov, *et al.* FEBS Lett. 277:230-232, 1990; Galyov, *et al.* 286:79-82, 1991; Karlyshev *et al.* FEBS Lett. 305:37-40, 1992). The role of these proteins in plague pathogenesis has not been unequivocally determined, and the effect of mutational loss of these proteins on the LD_{50} varies, depending on the animal model and route of infection (Brubaker Curr. Top. Microbiol. 57:111-118, 1972; Brubaker Rev. Infect. Diis. 5:S748-S758, 1983). However, pMT1 does appear to contribute to the acute phase of plague infection, as evidenced by a reduced morbidity associated with infection by strains lacking pMT1 (Drozdov, *et al.* J. Med. Microbiol. 42:264-268, 1995; Samoilova, *et al.* J. Med. Microbiol. 45:440-444, 1996;Welkos, *et al.* Contrib. Microbiol. Immunol. 13:229-305, 1995).

Information pertaining to the genetic characterization of the pMT1 molecule is limited. The size of the plasmid has been found to vary, either from variations in the versions of the plasmids or in technique to measure the plasmids, from 90 kb to 288 kb (Filippov, *et al.* FEMS Microbiol. Lett. 67:45-48, 1990). It is known that pMT1 is an integrative plasmid capable of integrating into *Y.*

pestis chromosome with high frequency and at multiple sites, with integration likely resulting from IS*100* homology between the plasmid and chromosome (Protsenko, *et al.* Microbiol. Pathogen 11:123-128, 1991).

Previous characterization of pMT1 has identified five genes that may be involved in the synthesis of murine toxin (MT) and F1 capsule antigen, both known virulence factors. Expression of both the capsular protein and murine toxin genes has been characterized with respect to environmental cues (e.g., temperature and calcium) (Du, *et al.* Contrib. Microbial. Immunol. 13:321-324, 1995). F1 capsule synthesis is maximal at 37°C in the absence of extracellular calcium, conditions similar to those that induce expression of a major *Y. pestis* virulence determinant (Straley Rev. Infect. Dis. 10:S323-S326, 1988; Straley Microbial. Pathogen 10:87-89, 1991; Straley *et al.* Proc. Natl. Acad. Sci. USA 78:1224-1228, 1981). Murine toxin expression is induced at 26°C, conditions similar to those that would be expected to occur in the flea vector. The occurrence of plasmid genes that are induced under widely different conditions suggests regulation of *Y. pestis* virulence determinant expression by at least two networks.

The plasmid pCD1 is found in *Y. pestis*, as well as in certain other pathogenic Yersinia species, including *Y. pseudotuberculosis* and *Y. enterocolitica*. The plasmid encodes a complex virulence property called the low-Ca^{2+} response (LCR). The LCR was discovered in *Y. pestis* growing *in vitro*, where the bacteria respond to the absence of Ca^{2+} at 37°C by the strong expression and secretion of a virulence protein called V antigen, or LcrV. In certain media, expression of LcrV is accompanied by a response termed "restriction," in which the yersiniae undergo an orderly metabolic shutdown and cease growth. Under

LCR-inductive conditions, the transcription, translation, and secretion of a set of virulence proteins called Yops (for Yersinia outer proteins) is maximally induced. The operons encoding these and other similarly regulated operons on the LCR plasmid have been referred to as the LCR stimulon (LCRS). Millimolar concentrations of Ca^{2+} permit full growth at 37°C, reduced expression of LcrV and Yops, and essentially no secretion of these proteins. Under ambient temperature conditions outside a mammalian host, the Yops and LcrV proteins are produced at a low, basal level and are not secreted, which suggests that the LCR is designed to function within a mammal. Expression of LCR is apparently modulated by other environmental factors, including Mg^{2+}, Cl^-, Na^+, glutamate, nucleotides, and anaerobiosity. The molecular basis for these effects has not been determined, but these elements of environmental modulation could be important in adjusting virulence protein expression and secretion in response to the wide range of niches that yersiniae are expected to encounter during an infection.

The pCD1 plasmid also encodes a type III secretion system called Ysc (for Yop secretion) that is involved in the secretion of Yops, LcrV, and some regulatory proteins in the LCR. The Ysc system is locally activated by cell contact at the interface between a bacterium and eukaryotic cell. This cell to cell contact causes the opening of the secretion system's inner and outer gates (LcrG and LcrE (or YopN), respectively), thereby allowing secretion of negative regulatory proteins (e.g., LcrQ also called YscM, a key regulatory protein). Secretion of negative regulatory proteins allows full transcriptional activation of LCRS operons by LcrF, an AraC-like activator protein.

Yops are secreted locally, without processing. The secretion mechanism recognizes two signals: one in the first 45 nucleotides of the yop mRNA and one related to a

domain that has been found for some Yops to bind a specific Yop chaperone (Syc), also encoded by the LCR plasmid . Certain of the Yops (e.g., YopB, YopD, YopK) are involved in targeting effector Yops (YopE, YopH, YpkA, YopM, and possibly YopJ) into the eukaryotic cell. Once inside the cell, the effector Yops act on intracellular target molecules, thereby interfering with cellular signaling and cytoskeletal functions. LcrV acts functions both as a regulatory protein involved in Yop secretion and targeting and as a potent anti-host protein. LcrV is the only LCRS protein that is secreted in large amounts into the surrounding medium by yersiniae in contact with eukaryotic cells. LcrV adversely affects the host organism when administered alone to mice, whereas all other secreted proteins depend on the Ysc machinery of yersiniae, in intimate contact with mammalian cells, for delivery into the mammalian cells.

Expression of the LCR has a profound immunosuppressive effect that results from the interference with innate defenses at the site of infection and the host organism's inability to mobilize an effective cell-mediated immune response. Y. pestis, and, in immunocompromised individuals, the enteropathogenic yersiniae grow unchecked in the lymphoid system in a fulminant disease associated with high mortality, absent appropriate antibiotic treatment. In contrast, yersiniae lacking the LCR plasmid pCD1 are completely avirulent.

Several other important pathogens have virulence systems with many striking similarities to the LCR; however, the LCR is the best characterized of these and remains a prototype for investigations at the forefront of molecular pathogenesis.

A more complete understanding of the role of LCR plasmids may be obtained by determining the entire sequence of an LCR plasmid.

The development of additional sequence information from plasmids of *Y. pestis* is needed for comprehensive efforts in the detection, diagnosis, prophylaxis and treatment of infections caused by the organism.

BRIEF SUMMARY OF THE INVENTION

One aspect of the present invention is an isolated *Yersinia pestis* plasmid pMT1- or pCD1-specific polynucleotide sequence selected from the group consisting of any portions of the sequences present in SEQ ID NO:1 through SEQ ID NO:6 set forth below.

The present invention is in part summarized by the presentation of the complete nucleotide sequence of two plasmids from *Yersinis pestis*, which enables diagnostic, prophylactic and therapeutic tools to be developed for use in combating the pathogen.

The DNA sequences of the present invention may include an open reading frame (ORF), an insertion sequence element, or a plasmid maintenance function, for example.

It is an object of the invention to provide essentially the entire sequence of pMT1 and pCD1 from *Yersinia pestis* KIM5 to allow methods of detecting, diagnosing, preventing, and treating infections with *Yersinia pestis*.

Other object, advantages and features of the present invention will become apparent from the following specification when taken in conjunction with the following drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Fig. 1 is a plasmid map of the plasmid pMT1, showing in schematic fashion the relative positions of notable features of the plasmid.

Fig. 2 is a similar plasmid map of the plasmid pCD1.

DETAILED DESCRIPTION OF THE INVENTION

This specification describes the complete DNA sequencing of the plasmids , pPCP1, pMT1 and pCD1 from *Yersinia pestis*, all of which are associated with the pathogenicity of the organism. Presented below is both the complete DNA sequence of the plasmids as well as tables listing the open reading frames (ORFs) of the plasmids, indicating which portions of the plasmid DNA encodes the production of proteins. Some other important regions of the plasmid DNA, such as the integration sequences (IS) are also indicated. With the information provided by this complete DNA sequence information, several things become possible. It now becomes possible to design and implement nucleotide-sequence based diagnostic tools to diagnose and identify virulent strains of *Yersinia pestis* in a biological sample based on the presence of DNA sequence in such a sample. The identification of the ORFs contained in the plasmids makes possible the comprehensive identification and characterization of the toxins and other proteins encoded by the plasmids thereby enabling the ability to make antibody and other molecular forms of prophylactic and therapeutic treatment for the pathogen. This information also allows identification of new potential virulence factors that may be useful in the development of vaccines, or which may be suitable targets for therapeutic drugs. In addition, the sequencing data provides information about maintenance functions, horizontal gene transfer, conjugation, integration, insertion sequence (IS) elements, and evolution of these plasmids. The sequences from pCD1 and pMT1, and their significance, were first published by the inventors here in Lindler *et al.* Inf. Immunity 66:5731-5742, 1998 and Perry *et al.* Inf. Immunity 66:4611-4623, 1998, both of which are incorporated herein by reference in their entirety. Identification of maintenance functions provides

information that is useful in designing cloning vectors, which can be used, for example, to study factors associated with pathogenicity.

Briefly, as described below in the examples, we determined the entire nucleotide sequence of the plasmid pMT1 from *Y. pestis* strain KIM5. We then analyzed the sequence and identified potential open reading frames (ORFs) encoded by the 100,990 bp pMT1 molecule. The complete sequence is set contained in SEQ ID NO 2 below. Based on yersinial codon usage for known yersinial genes, homology with known proteins in the databases and potential ribosome binding sites, it was determined that 115 of the potential ORFs likely encode proteins in *Y. pestis*. Seven new potential virulence factors that might interact with the mammalian host or flea vector were identified. The deduced amino acid sequences for 43 of the remaining 115 putative ORFs display no significant homology to proteins in the current databases. Furthermore, DNA sequence analysis allowed the determination of the putative replication and partitioning regions of pMT1.

A single 2,450 bp region within pMT1 that may function as the origin of replication (ori) was identified. The identification of this putative ori may allow construction of cloning vectors capable of replicating in Yersinia species. Such vectors will facilitate further research into the pathogenicity of these bacteria. The putative ori includes a RepA-like protein similar to those of the RepFIB, RepHI1B, P1 and P7 replicons. A plasmid partitioning function is located about 36 kilobases from the putative origin of replication and is most similar to the *parABS* bacteriophage P1 and P7 system. *Y. pestis* pMT1 encodes potential genes with a high degree of similarity to a wide variety of organisms, plasmids and bacteriophage. Accordingly, our analysis of pMT1 DNA sequence suggests the mosaic nature of this large bacterial virulence plasmid and

provides insight into its evolution. The MT- and F1 encoding regions of pMT1 are surrounded by remnants of multiple transposition events and bacteriophage, respectively, suggesting horizontal gene transfer of these virulence factors.

The pCD1 sequence is 70,509 base pairs, and is presented as SEQ ID NO:1 herein. The SEQ ID NO:1 is actually 70,559 base pairs in length since it incorporates a 50 base pair repeat at each end of the linear representation of the circular plasmid. Sequencing of pCD1 has revealed a potential new Yop and Yop chaperone, two new IS, a set of LCRS genes very similar to those sequenced in the enteropathogenic yersinae, the IncFIIA replication region, and SopABC partitioning functions. Remnants of IS elements were found to be scattered throughout the plasmid, which suggests that pCD1 has undergone numerous insertional events as well as genetic recombinations and rearrangements during its history.

Yersinia pestis has an unique 9.5-kb plasmid, designated pPCP1, which contains genes encoding plasminogen activator/coagulase and pesticin. The total length of pPCP1 is 9,610 bp with a GC of 43%. The plasmid pPCP1 contains a copy of IS100. Three known gene functions located on this plasmid are as follows: 1) plasminogen activator and coagulase activity that is encoded on the same gene (pla), 2) pesticin, a toxin that inhibits growth of closely related bacteria, and 3) pesticin immunity gene whose product protects the bacteria from toxic effects of the pesticin. The origin of replication of pPCP1 is encoded on 780 bp region which is very similar to the origin of replication and the immunity region of Escherichia coli ColE1 plasmid. Loss of this plasmid leads to ineffective infection in guinea pigs and mice suggesting that the plasmid plays an important role in the invasion and infection of its mammalian host. The plasmid pPCP1 has also

been sequenced and its sequence is presented as SEQ ID NO:3.

The sequences presented here are accurate to the best capabilities of the current state of the art, but may contain some minor errors, deletions, insertions or substitutions. It is also understood and expected that other strains of the host organism will have allelic variations of the genes in the host and therefore may carry different forms of the genes set forth in the sequence listing here. However, those of skill in the art expect such minor variations, and such minor sequence variations in *Yersinia pestis* -specific nucleotide sequences associated with nucleotide additions, deletions, and mutations, whether naturally occurring or introduced *in vitro*, would not interfere with the usefulness of these sequences in the detection of *Yersinia pestis*, in preventing *Yersinia* infection, and in methods for treating *Yersinia pestis* infection. Therefore, the scope of the present invention is intended to encompass such variations in the claimed sequences.

A *Yersinia pestis* -specific nucleotide probe is a sequence that is able to hybridize to *Yersinia pestis* target DNA present in a sample containing *Yersinia pestis* under suitable hybridization conditions, and which does not hybridize with DNA from other *Yersinia species* or from other bacterial species. It is well within the ability of one skilled in the art to determine suitable hybridization conditions based on probe length, G+C content, and the degree of stringency required for a particular application.

The probe may be RNA or DNA. Depending on the detection means employed, the probe may be unlabeled, radiolabeled, or labeled with a dye. The probe may be hybridized with a sample that has been immobilized on a solid support such as nitrocellulose or a nylon membrane, or the probe may be immobilized on a solid support, such as

a silicon chip.

The sample to be tested for presence or absence of *Yersinia pestis* DNA may include blood, urine, feces, or other materials from a human, rodent, or flea susceptible to infection by yersinia pestis. The sample may be tested directly, or may be treated in some manner prior to testing. For example, the sample may be subjected to PCR amplification using appropriate oligonucleotide primers. To have reasonable assurance of success under conditions of variable stringency, it is preferred that such diagnostic probes uses sequences which are at least 15 nucleotides or longer in length. While probes as short as 15 base pairs can be made to work, probes of at least 25 base pairs or longer are preferred. Any means of detecting DNA-RNA or DNA-DNA hybridization known to the art may be used in the present invention. Since the plasmids set forth below are diagnostic of pathogen strains of *Yersinia pestis*, any set of 25-mers or longer from the sequences set forth below may usefully be employed as diagnostic probes for the presence of this pathogen in a biological sample.

Any and all of the ORFs presented here are of particular utility. Since these ORFs contain the coding regions for the proteins expressed by these plasmids, these ORFs are not just useful for diagnosis of the presence of the pathogenic host, they may be used to express the encoded proteins in other hosts. Placing the coding regions of the ORFs under the control of non-native promoters permits the expression of the proteins encoded by the ORFs in other hosts. The ORFs can be inserted into any known expression vector adapted for a particular host and then can be transformed into that host for expression to produce proteins. Such proteins can be used for both prophylactic and therapeutic purposes. The proteins can be used to generate antibodies to the proteins natively produced by the *Y. pestis*, the provide pathogen specific

antibodies for diagnostic or therapeutic purposes. Proteins, or even peptides from the proteins have potential for targets for vaccination studies.

EXAMPLES

Isolation of pMT1 DNA.

Y. pestis KIM10+ (Perry, et al. J. Bacteriol. 172:5929-5937, 1990), a strain that contains only pMT1, was grown in Heart Infusion Broth (Difco Laboratories, Detroit, Michigan) at 26-30°C. Plasmid DNA was isolated from the bacteria using alkaline lysis and polyethylene glycol precipitation (Birnboim, et al. Nucleic Acids Res. 7:1513-1523, 1979; Humphreys, et al. Biochim. Biophys. Acta 383:457-463, 1975). DNA libraries were prepared from purified pMT1, as described below.

Isolation of pCD1 DNA.

Y. pestis strain KIM5 is conditionally avirulent due to deletion of the 102 kb pgm locus; it possesses all three prototypical *Y. pestis* plasmids (pPCP1, pCD1, and pMT1). Plasmid DNA was isolated from *Y. pestis* KIM5 by alkaline lysis followed by precipitation with polyethylene glycol. A mixture of pCD1 and pBR322 was transformed into *Escherichia coli* HB101. Transformants containing pBR322 were selected on the basis of ampicillin resistance. Ampicillin resistant transformants were transferred to nitrocellulose membranes and hybridized against pCD1 radioactively-labeled by nick translation, which allowed identification of cotransformants containing both pCD1 and pBR322. A selected cotransformant was cured of pBR322 by fusaric acid selection and used for isolation of pCD1. The pCD1 plasmid appears to be stably maintained in *E. coli* HB101. Plasmid DNA from *E. coli* HB101 (pCD1) cells grown in Luria broth was isolated by alkaline lysis followed by

further purification with polyethylene glycol. Purified pCD1 DNA was used in subsequent sequencing.

pPCP1

DNA of pPCP1 was isolated for sequencing in a similar fashion.

DNA sequencing.

DNA libraries of pPCP1, pCD1 or MT1 were prepared from nebulized, size fractionated plasmid DNA (Millon, et al. Gene, submitted) in the M13 Janus vector (Burland, et al. Nucleic Acids Res. 21:3385-3390, 1995). DNA templates were purified from random library clones (Romantschuk, et al. Mol. Microbiol. 5:617-622, 1991), and DNA sequencing was preformed using dye-terminator labeled fluorescent cycle sequencing Prism reagents and ABI377 automated sequencers (Applied Biosystem Division of Perkin-Elmer). Sequences were assembled into segments of DNA sequence, referred to as contigs, by the SeqMan II program (DNASTAR), and clones were selected for sequencing from the opposite end to fill in coverage, resolve ambiguities and close gaps. Final coverage was about eight fold. The complete sequences of ball three plasmids are set forth in SEQ ID NO: 1 through 3 below.

In several instances, pCD1 sequences differed from previously published sequences from the yersiniae or yielded unexpected results. To ensure that this did not result from mutations to pCD1 during carriage in *E. coli*, we sequenced these regions using pCD1 isolated from the conditionally virulent *Y. pestis* strain KIM5 or pJIT7, a recombinant plasmid containing the IS1616 region adjacent to sopAB.

Sequence Annotation.

Open reading frames (ORFs) putatively encoding polypeptides at least 50 aa in length were identified using Geneplot or GeneQuest (DNASTAR) programs to display start

codons (including GUG), stop codons and codon usage statistics plots for each reading frame. Codon usage analysis, used to predict ORFs, was assessed in the program by second and third order statistical comparisons with a matrix built from all available sequences for Yersinia species (Borodovsky, et al. Computational Chemistry 17:123-133, 1993). Although this matrix was more useful than one derived from *E. coli* genes, it was necessarily constructed from a relatively small data set. Generally, the start codon (including GTG and TTG) farthest upstream was used to annotate the ORF start. An ORF having fewer than 150 bases was included if it had a high codon usage score. For the first pass, putative amino acid sequences were searched against SWISS-PROT 34 using the BLOSUM26 matrix, by the DeCypher II System (TimeLogic Inc., Incline Village, Nevada).

Subsequent searches of the Swiss Protein, *E. coli* and non-redundant GenBank databases were obtained over the Internet using BLAST software (Altschul, et al., Nucelic Acids Res. 25:3389-3402, 1997) from the National Center for Biotechnology Information homepage (www.ncbi.nlm.gov/BLAST/). Pairwise protein alignments were with the BLAST algorithm. Protein localization was predicted for relevant translated *orfs* using the PSORT program (Nakai, et al. Proteins: Structure, Function, and Genetics 11:95-110, 1991). The prediction of membrane associated helices was with the TMpred program (Hoffman, et al. Biol. Chem. 347:166-172, 1993). Where appropriate, multiple protein sequences were aligned using the algorithm developed by Lipman *et. al.* (Proc. Natl. Acad. Sci. USA 86:4412-4415, 1989). These programs can be found as part of Pedros Molecular Biology Tools at Internet site www.iastate.edu.

Bank accession number.

The annotated sequence for pMT1 and pCD1 were deposited in GenBank under accession numbers AF074611 and AF074612, respectively. These deposited sequences are also hereby incorporated by reference.

Sequence of pMT1

The fully-assembled pMT1 DNA sequence is a circular DNA sequence 100,990 bp in length. A map of the plasmid is set forth in Fig. 1, which illustrated the general location of sequences of interest. The complete DNA sequence of the plasmid is presented here as SEQ ID NO:2. Screening of the entire plasmid sequence using the DNASTAR program GeneQuest revealed 145 potential open reading frames (ORFs) along the entire length of the plasmid. The putative amino acid sequence of each ORF was used to search the various databases (GenBank, Swiss Protein,GenPept and *E. coli*) for proteins with potentially significant homologies. Table 1, set forth below, identified the location and other information of interest about many of the ORFs which were found to have homologies to known sequences.

Table 1. ORFs identified in *Y. pestis* pMT1 DNA sequence by classification.[a]

Designation	ORF Class	Function or Comments	Organism or Element (Gene if known)	Accession Number	Location (bp)
DNA Metabolism					
	ORF1	IS*100*	*Y. pestis* IS*100* (*orfB*)	U59875	73,885-74,661
	ORF2	Ligase	Bacteriophage T3	X05031	74,680-75,777
	ORF12	Integrase	*Vibrio cholera*	U39068	82,931-84,109
	ORF16	DNA Pol III	*E. coli*	M19334	88,955-92,479
	ORF26	RecA	*Bacteroides fragilis* (recA)	M63029	96,910-97,986
	ORF34	RepA	*E. coli* plasmid ColV	L01250	Complement 717-1,781
	ORF41	*exoA*	Bacteriophage T4 (gene 47)	X01804	4,968-6,053
	ORF43	*exoB*	Bacteriophage T4 (gene 46)	X01804	6,271-8,199
	ORF46	IS*200*	IS*200*	U22457	9,675-10,184
	ORF60	Rep-like	*Coxiella burnetti* plasmid pQPH1	L34077	16,197-16,895
	ORF61	SpoJ-like	*Streptococcus pneumoniae*	AF000658	16,862-17,563
	ORF69	Gene 17-like	Bacteriophage T4 (gene 17)	X52394	20.457-21,713
	ORF93	IS*100*	*Y. pestis* IS*100 (orfB)*	U59875	Complement 46,449-47,231

	ORF94	IS*100*	*Y. pestis* IS*100 (orfA)*	U59875	Complement 47,228-48,250
	ORF101	IS*285*	*Y. pestis* IS*285 (orf2)*	X78303	51,013-52,221
	ORF102	Transposase TN4321	Enterobacter aerogenasesTN4321(tnpA)	U60777	52,648-53,712
	ORF108	Membrane Endonuclease	*E. coli* plasmid pKM101 (*nuc*)	U09868	Complement 57,629-58,117
	ORF111	Resolvase	*Pseudomonas syringae (stbA)*	L48985	Complement 60,161-60,781
	ORF113	ParA	Bacteriophage P1 (parA)	X02954	61,767-63,041
	ORF114	ParB	Bacteriophage P1 (parB)	K02380	63,038-64,009
	ORF123	Adenine specific DNA methylase	*E. coli* pEC156 EcoVIII methylase	U48806	66,648-67,325
	ORF128	Antirestriction	*E. coli*	Z34467	69,208-69,714
	ORF135	DNA Partitioning	*Rhizobium meliloti* (Orf1, Orf2 of pRmeGR4a), *Shigella sonnei (psiB)*, *Streptoccus pneumoniae (spoOJ)*	X69105, U82272, AF000658	70,730-72,739
	ORF136	IS*100*	*Y. pestis* IS*100 (orfA)*	U59875	72,863-73,882
Protein Metabiolism					
	ORF28	HflC-like	*Vibrio parahaemolyticus* (hflC)	U09005	98,281-99,111

	ORF63	ABC transporter/ATP binding	*Archaeglobus fulgidus* (AF1064)	AE001029	17,500-18,198
	ORF75	L12 Ribosomal protein L12e	*Haloferax volcanii*	X58924	25,927-26,361
Gene Regulation					
	ORF5	Caf1R	*Y. pestis (caf1R)*	X61996	Complement 77,118-78,041
	ORF22	PprB-like	*Pseudomonas putida* (pprB)	X80272	94,557-95,636
	ORF56	Repressor of flagella synthesis	*Salmonella abony (fljA)*	D26167	Complement 13,278-13,841
Known Virulence					
	ORF6	Caf1M	*Y. pestis (caf1M)*	X61996	78,318-79,127 (GTG Start)
	ORF8	Caf1A	*Y. pestis (caf1A)*	X61996	79,152-81,653
	ORF9	Caf1	*Y. pestis (caf1)*	X61996	81,734-82,246
	ORF107	Murine toxin	*Y. pestis (ymt)*	X92727	Complement 55,788-57,551
Lambda-like					
	ORF80a	V major tail fiber Intimin	Bacteriophage lambda E. coli O157:H7 *(eae)*	P03733 P43261	28,560-29,303
	ORF84	H tail fiber protein	Bacteriophage lambda	AF007380	30,041-34,618

	ORF85	M minor tail fiber protein	Bacteriophage lambda	P03737	34,660-34,995
	ORF86	L minor tail fiber protein	Bacteriophage lambda	P03738	35,052-35,783
	ORF87a	K tail assembly protein	Bacteriophage lambda	P03729	35,815-36,570
	ORF88	I tail assembly protein	Bacteriophage lambda	P03730	36,561-37,148 (GTG Start)
	ORF89	J host specificity protein	Bacteriophage lambda	P03749	37,164-41,801
	ORF91	Hypothetical protein ORF314	Bacteriophage lambda	P03745	42,469-45,405
	ORF92	Tail fiber assembly	Bacteriophage lambda *(tfa)*	225931	45,707-46,315
Hypothetical in database[b]					
	ORF15	CobT	*Pseudomonas denitrificans (cobT)*	P29934	85,075-87,441
	ORF15a	CobS	*Pseudomonas denitrificans (cobS)*	P29933	87,539-88,771
	ORF29	Hypothetical protein	Bacteriophage P22 (ninX)	X78401	99,265-99,636
	ORF33a	Hypothetical regulatory protein	Bacteriophage P1	76816	100,922-147
	ORF38	Hypothetical lipoprotein	*Bacillus subtilis (orfK, yzeA)*	L16808, Z93102	Complement 3,530-4,552
	ORF59	Long hypothetical protein	*Pyrococcus horikoshii* (PHBW005)	AB009472	Complement 14,573-16,132

	ORF73	SRPI Hypothetical protein	*Synechococcus* PCC7942 pANL	Q55032	24,271-25,146
	ORF104	Hypothetical protein	*E. coli*	U70214	Complement 54,408-54,803
	ORF105	Hypothetical protein	*E. coli*	U70214	Complement 54,694-55,002
	ORF116	Hypothetical protein	*Sphingomonas* S88 (spsJ)	U51197	64,388-65,785
	ORF131	Hypothetical protein	*E. coli*	AE000133	70,427-70,657
Fragments[c]					
	ORF23	DNA polymerase I	*Lactococcus lactis*	U78771	95,646-96,641
	ORF33	Type II restriction enzyme	*Helicobacter pylori*	AE000647	100,590-100,925
	ORF99	Hypothetical protein	*Methanobacterium thermoautotrophicum*	AE000913	Complement 49,210-50,004
	ORF103	Hypothetical transposase	*Salmonella typhimurium*	Z29513	Complement 53,911-54,234
	ORF103a	IS*600*	*Shigella sonnei*	X05952	54,281-54,481
	ORF106	Hypothetical	*Shigella flexneri*	U97489	55,073-55,543
	ORF106a	IS*801*	*Pseudomonas syringae*	X57269	55,589-55,729
	ORF110	Hypothetical	*Salmonella typhimurium*	Z29513	Complement 59,154-60,140
	ORF115a	SamB-like	*Salmonella typhimurium*	D90202	87,539-88,771

In the above Table 1, the location of each of the ORFs is given in base pair number corresponding to the entire 100,990 base pairs of the entire plasmid. ORFs listed were assigned a putative function according to our criteria outlined in the general overview section of the results and discussion. Classification then was based on these putative functions.

If there was insufficient homology, by our criteria, with known proteins in the database the ORF has not been assigned a function in the table. In evaluating the significance of potential matches, several factors were considered. In general, if the putative translation product of a pMT1 ORF exhibits significant similarity to known proteins in the database, the putative protein was assigned a similar function. Homologies were considered to be significant if at least 25 percent of amino acids were identical over at least 35 percent of the protein in the database. The 25% identity was chosen to give a reasonable baseline, with adjustments being made for conservative amino acid substitutions to give higher similarity scores between protein molecules.

In specific instances, we have designated a protein function as "similar" based on less than 25 percent identity. The extent of homology with the database protein was set at 35 percent to allow for the possibility that protein domains might have different functions in different molecular contexts. The stringency was lowered when deciding if a putative protein might function in pathogenesis. In these cases, if the region of homology included at least 20 percent identical amino acids with a protein that might interact with or substitute for the action of a host protein, it was considered a potential virulence factor. Greater weight was given to potential alignments if the homology between the *Y. pestis* ORF and the target protein sequence was in a domain having a known function in host physiology. Finally, if the

putative protein does not contain significant similarity to any known proteins, the upstream DNA was analyzed for ribosome binding sites (RBS) and the known codon usage for *Yersinia* genes was considered. After applying these criteria to the 145 potential ORFs initially identified on pMT1, 30 were eliminated and 115 putative coding regions remained. Of these 115 putative ORFs, 38 percent had no significant regions of homology to any protein in the current databases and seven percent had significant homology with previously described hypothetical proteins.

Newly identified virulence factors of pMT1.

Because *Y. pestis* is a facultative intracellular parasite and pMT1 is thought to enhance deep tissue spread of the organism, several ORFs having limited homology with proteins that may function during various stages of the plague life cycle were carefully examined. The ORFs include ORF 4 (base pairs 76,298 to 76,603), ORF 17 (bases 92,476-92,919), ORF 18 (complement to bases 92,949-93,512), ORF 21 (bases 94,015-94,448), ORF 72 (23,873-24,244), and ORF 74a (25,221-25,883). Again, all base pairs locations refer to the complete 100,990 sequence. Additional information about these identified virulence factors is presented in Table 2, below. Although many of these homologies are below our criteria for general ORF homologies, a more relaxed standard was indicated to aid in future research relating to plague pathogenesis.

Table 2. ORFs that may be potential virulence factors.

ORF Designation	Location	Homologus Protein (Target)	Amount of Homology [a]	Accession Number	Reference
ORF4	76,298-76,603	C-type natriuretic peptide from *Squalus acanthias*	43/30	P41319	83
ORF17	92,476-92,919	Delta insecticidal protein from *Bacillus thuringiensis*	40/18	P05628	35
ORF18	Complement 92,949-93,512	RTX Toxin of *Actinobacillus pleuropneumoniae*	21/11	D16582	32, 65
ORF21	94,015-94,448	Laminin of *Homo sapiens,*	23/5	Q16787	79, 95
		Paramysin-related protein of *Onchocerca gibsoni*	21/18	U20609	25, 99
ORF72	23,873-24,244	Major Myristoylated Alanine-rich Protein Kinase C Substrate (MARCKS)	24/32	P29966	41
ORF74a	25,221-25,883	Bacteriophage lambda V protein,	40/41	P03733	81
		Citrobacter freundii intimin	30/10	Q07591	82

a. Percent identical amino acids over the percent of the total target protein sequence.

In addition, one potential new IS element, designated IS*1618*, is located from bp 52,465 to 53,758 (or bases 2365-3658). This sequence, the boundaries of which are defined by

two directly repeated sequences (GATGATAA), flanks a putative transposase designated ORF102. ORF102 had the greatest identity with a putative transposase previously found in *Enterobacter aerogenases* (Smith, et al. J. Gen. Microbiol. 139:1761-1766, 1993)(40% over 96 percent of the target protein) and a putative transposase previously described in *Yersinia enterocolitica* (Rakin et al. FEMS Microbiol. Lett. 129:287-292, 1995) (36% identity over 96% of the target protein).

The nucleotide sequence of *Y. pestis* pMT1 has provided a wealth of new information. Our analysis has allowed us to identify several genes to target for further study in order to access their possible role in pathogenesis. Deciphering the potential role of these proteins improves our understanding of disease as well as host physiology. As more complete virulence plasmid DNA sequences become available, we will begin to understand the mosaic nature of these molecules and what new combinations we might expect in the future. Detailed molecular analysis of the structure of virulence plasmids will impact our ability to predict the emergence of bacterial pathogens as well as detect their presence.

Sequences of pCD1

A genetic map of the Y. pestis KIM5 pCD1 plasmid, which is 70,509 nucleotides in length, is shown in Fig. 2. Again the complete DNA sequence of the plasmid is contained in the sequence listing appended hereto, this sequence being SEQ ID NO:1. Again the ORFs of the sequence was determined by computer analysis and searched against existing data bases. Table 3 below lists significant ORFs and their primary characteristics. Most IS element remnants and partial ORFs that appear to be nonfunctional due to IS-related events or other deletions and rearrangements are not included in Table 3.

Table 2. ORFs encoded on pCD1 of Y. pestis KIM5a

geneb or ORF	Function	Orienta -tion	Begin -ing of ORF	End of ORF	Number of amino acids	Isoelec -tric point	kDa
repB (copB)	Negative regulator of repA transcription	+	1,171	1,425	85	9.72	9.58
tap	Required for translation of repA	+	1,667	1,741	25	9.31	2.82
repA	Plasmid replication	+	1,734	2,600	289	10.96	33.55
Orf5	Unknown	-	3,645	3,427	73	9.96	8.22
Orf7	Unknown	+	4,758	5,186	143	4.39	15.78
ypkA (yopO)	Targeted effector; ser thr kinase	+	5,204	7,402	733	6.53	81.74
yopJ (yopP)	Targeted effector; causes apoptosis in macrophages and interferes with cell signaling	+	7,798	8,664	289	7.07	32.46
yopH	Targeted effector; protein tyrosine kinase; interferes with cell signaling at focal adhesions	+	10,347	11,753	469	8.68	50.87

lcrQ (yscM)	Negative regulator of LCR expression	-	16,148	15,801	116	6.34	12.41
yscL	Type III secretion component	-	17,038	16,373	222	4.57	24.65
yscK	Type III secretion component	-	17,613	16,984	210	6.75	23.99
yscJ	Type III secretion component	-	18,347	17,613	245	7.43	27.04
yscI	Type III secretion component	-	18,701	18,354	116	4.47	12.67
YscH (yopR)	Secreted; unknown function	-	19,199	18,702	166	5.14	18.35
yscG	Type III secretion component	-	19,543	19,196	116	6.60	13.07
yscF	Type III secretion component	-	19,808	19,545	88	7.13	9.49
yscE	Type III secretion component	-	20,009	19,809	67	7.31	7.61
yscD	Type III secretion component	-	21,265	20,006	420	5.85	46.93
yscC	Type III secretion component	-	23,085	21,262	608	6.49	67.35

yscB	Unknown	-	23,504	23,091	138	9.27	15.41
yscA	Unkown	-	23,828	23,730	33	9.82	3.86
lcrF (virF)	Activator or LCR expression	-	24,722	23,907	272	8.91	30.84
yscW (virG)	YscC lipoprotein chaperone	-	25,241	24,846	132	10.12	14.71
geneb or ORF	Function	Orienta-tion	Begin-ing of ORF	End of ORF	Number of amino acids	Isoelec-tric point	kDa
yscU	Type III secretion component	-	26,881	25,817	355	8.81	40.39
yscT	Type III secretion component	-	27,666	26,881	262	5.67	28.45
yscS	Type III secretion component	-	27,929	27,663	89	6.32	9.57
yscR	Type III secretion component	-	28,584	27,931	218	4.68	24.43
yscQ	Type III secretion component	-	29,504	28,581	308	5.08	34.42
yscP	Type III secretion component	-	30,868	29,501	456	5.44	50.42
yscO	Type III secretion component	-	31,332	30,868	155	7.84	19.00

yscN	Type III secretion component	-	32,648	31,329	440	6.48	47.81
lcrE (yopN)	Secretion control	+	32,846	33,727	294	5.07	32.67
tyeA	Secretion and Yop targeting control	+	33,708	33,986	93	4.21	10.75
Orf42	Unknown	+	33,973	34,344	124	5.54	13.61
Orf43	Unknown	+	34,341	34,709	123	6.32	13.76
Orf44	Unknown	+	34,706	35,050	115	6.92	13.12
lcrD (yscV)	Secretion	+	35,037	37,151	705	5.04	77.81
lcrR	Unknown	+	37,148	37,588	147	10.27	16.46
lcrG	Secretion control; efficient Yop targeting	+	37,630	37,917	96	8.15	11.02
lcrV	Diffusible effector; secretion and targeting control	+	37,919	38,899	327	5.66	37.24
lcrH (sycD)	YopB and YopD chaperone	+	38,912	39,418	169	4.61	19.02
yopB	Yop targeting	+	39,396	40,601	402	7.09	41.83

yopD	Yop targeting; negative regulator	+	40,620	41,540	307	6.80	33.39
Orf54	Unknown	-	42,709	42,386	108	9.66	12.61
yopM	Targeted effector	+	43,481	44,710	410	4.23	46.21
Orf60	Unknown	-	46,365	45,946	140	7.79	15.81
Orf61	Unknown	+	46,637	47,026	130	7.33	14.80
sycT	YopT chaperone	-	47,468	47,070	133	4.43	15.42
yopT	Targeted effector	-	48,436	47,468	323	9.13	36.31
yopK (yopQ)	Yop targeting	+	48,936	49,484	183	4.37	21.00
ylpA	pseudogene	+	50,089	50,718	210	5.80	22.40

geneb or ORF	Function	Orienta -tion	Begin- ing of ORF	End of ORF	Number of amino acids	Isoelec -tric point	kDa
sopA	Plasmid partitioning; negative regulator of sopAB transcription	+	52,730	53,896	389	5.82	43.41
sopB	Plasmid partitioning; binds to sopC region	+	53,896	54,858	320	10.19	35.61

Orf73	Unknown	+	56,087	56,362	92	6.16	10.10
Orf74	Unknown	+	56,355	56,654	100	5.51	11.67
Orf75	Unknown	-	56,792	56,496	99	9.88	11.19
yopE	Targeted effector; causes actin depolymerization	-	57,453	56,794	220	6.59	22.99
sycE (yerA)	YopE chaperone	+	57,647	58,039	131	4.49	14.65
sycH	YopH chaperone	+	60,796	61,221	142	4.81	15.76
Orf84	Unknown	-	62,897	62,568	110	8.98	13.00
Orf85	Unknown	-	63,500	63,036	155	4.97	17.71
yadA'	pseudogene	+	67,532	67,783	84	5.21	8.92
'yadA	pseudogene	+	67,900	68,835	312	6.84	32.47

a ORFs within transposable elements as well as disrupted or partial ORFs (except for ylpA, yadA', and `yadA) are not included in the table.

b Except for copB, yopN, yscV, and yerA, all alternate gene designations, in parentheses, are Y. enterocolitica terminology; copB - plasmid R100 terminology; yopN - Y. enterocolitica and Y. pseudotuberculosis terminology; yscV - proposed terminology change; yerA - Y. pseudotuberculosis terminology

New potential virulence-related ORFs

Fourteen ORFs are not obviously associated with IS elements and either have no significant similarity to proteins in the database with known functions or have features suggesting a virulence-related role. These are ORFs that deserve future study as potentially having virulence or virulence-accessory functions.

ORF75 (Table 3) lies just 1 bp downstream of yopE and lacks

an obvious ribosome binding site or upstream promoter. The ORF could encode an 11,192 Da protein with at least one likely transmembrane domain and a noncleavable signal sequence. Its expression conceivably is translationally coupled to that of yopE suggesting that it could be a member of the LCR. yopE has been called monocistronic, based on its estimated transcript size (750 bases in *Y. pseudotuberculosis*). The presence of this ORF has not been noted in the literature, even though the beginning of Orf75 is present in the sequences previously submitted for *Y. pseudotuberculosis* yopE , *Y. enterocolitica* O:9 and *Y. pestis* EV76 . Interestingly, it is intact but separated from yopE by an insertion element in *Y. enterocolitica* O:8 strain 8081 . At high doses, a *Y. pseudotuberculosis* mutant containing an insertion in this ORF did not cause loss of virulence in mice infected orally (Forsberg, et al. J. Bacteriol. 172:1547-1555, 1990). Given that YopE's importance in virulence was determined with polar insertion mutants, the significance of this ORF needs to be thoroughly tested.

While assembling this data, we learned that two new ORFs we found in *Y. pestis* have been designated as YopT and SycT in *Y. enterocolitica* (Miller, *et al.* J. Bacteriol. 172:1062-1069, 1991). sycT and yopT are arranged in what appears to be a bicistronic operon upstream 500 bp and on the opposite strand from yopK (Fig. 2). These genes indeed have properties suggestive of a Yop and associated Syc. sycT is predicted to encode an acidic 15.42 kDa peripheral protein (Table 3). The database search brought up weak homology with SycE (with which there is 22% identity). Alignment of SycT with SycE, LcrH (SycD), and SycH shows the greatest similarity toward the C termini of the proteins, as previously demonstrated in a comparison of SycE and LcrH/SycD . YopT is predicted to be a peripheral 36.31 kDa basic protein (Table 3). It shows 36.7% identity in residues 98-322 with the C-terminus (residues 648-874) of a surface antigen in *Haemophilus somnus* that is associated with serum-resistance . The regulation, mechanism of

action, and role in plague of YopT should be investigated.

ORFs 42, 43 and 44 (Table 3), located immediately downstream of tyeA (Fig. 2), have been noted to exist in *Y. enterocolitica* (Winans *et al.* J. Bacteriol. 154:117-1125, 1083). ORF42 has been sequenced in *Y. pseudotuberculosis* and a polar insertion near its 3′ end caused a calcium-independent growth phenotype (Forsberg, et al. Mol. Microbiol. 2:121-133, 1988), typical of mutations in genes necessary for the functioning of the type III secretion system. Because this mutation was complemented by DNA lacking a complete lcrD/yscV gene (downstream of ORF44), the phenotype is not likely to be caused by disruption of lcrD/yscV. This, taken together with their location (within the LCR cluster and downstream of tyeA, which is involved in Yop secretion control), suggests that one or more of the ORFs 42 through 44 have a role in secretion or secretion control.

ORF5 (Table 3) is isolated from other virulence-related genes, within a gap between the origin region and an IS1236 remnant. It is presently unknown whether the sequence encodes a virulence-related factor.

ORFs 59, 60, and 61 (Table 3; Fig. 2) lie between yopM and sycT. Orf59 is closest to yopM (242 bp away), on the opposite strand, and is predicted to encode a 4 kDa soluble acidic protein (Table 3), which is significantly smaller than typical Sycs. Orfs 60 and 61 lie 875 bp from Orf59, are separated by 272 bp, and are divergently oriented. Both are predicted to encode membrane-associated proteins with mildly basic pIs that hence do not resemble typical Sycs (acidic, soluble, ca. 16 kDa) or Yops (soluble). Orf 60 has an uncommon translation initiation codon (leucine) (Table 3).

ORFs 73 and 74 (Table 3) lie in the vicinity of yopE. The predicted proteins are 10-11 kDa soluble acidic proteins that show high similarity to unknown proteins of similar lengths in *Mycobacterium tuberculosis;* however, neither ORF has a common translation initiation codon (leucine [ORF73] and valine

[ORF74]). Both ORFs are predicted to be transcribed in the same direction, with Orf74 overlapping Orf73 by 8 bp (Table 1).

ORFs 84 and 85 (Table 3; Fig. 2) occupy the region between IS1617 and Tn1000p. They are separated by 139 bp and would be transcribed in the same direction. The predicted product of Orf84 is a basic soluble protein and the product of Orf85 is predicted to be an acidic soluble protein (Table 3).

We identified a number of intact, defective, and partial IS elements in pCD1. The site of an IS100 insertion, an element with numerous copies in the *Y. pestis* genome (Fetherston, et al. Mol. Microbiol. 13:697-708, 1994; Portnoy, et al. Infect. Immun. 43:108-114, 1984), was confirmed and refined. Two new IS elements, which we have named IS1616 and IS1617, were discovered (Fig. 2) and were registered through Dr. Esther Lederberg Plasmid Reference Center, Stanford, CA. In addition, numerous IS element remnants were identified; these partial ISs primarily cluster in four regions of pCD1 (discussed below).

It is curious that IS100 is nearby one end of the yscM to yopD LCR cluster and two partial IS285 elements bound this same region (Fig. 2). The type III secretion system and regulatory genes, exemplified by this LCR cluster, is widespread among bacterial pathogens and has been suggested as a possible pathogenicity island (PAI) . PAI hallmarks include carriage of virulence genes, a distinct GC content compared to the host bacterium, a discrete genetic unit often flanked by direct repeats, association with tRNA genes and/or insertion sequences, presence of "mobility" genes (transposases, etc), instability, and absence in less pathogenic strains . An additional requirement of a chromosomal location may be somewhat artificial given the large sizes of many virulence plasmids. Although the LCR cluster does have IS elements associated with it, we failed to detect any tRNA genes anywhere on pCD1. In addition, the LCR cluster does not contain effector Yops (except for lcrV). Finally, the GC content of this region (44.8%) matches that of the entire plasmid

and is similar to the 46-47% GC content of the genome of Y. pestis.

Insertion elements. Several mobile genetic elements have been found in the pathogenic yersiniae and most of them are present on LCR plasmids as well as the chromosome . ISs known to be associated with the LCR plasmid of *Y. pestis* include IS100 and IS285 . Additional elements are found on the LCR plasmid of *Y. enterocolitica* but are not present on the *Y. pestis* plasmid . Sequence analysis of pCD1 from *Y. pestis* KIM5 revealed the presence of three complete insertion elements and numerous partial IS elements. Complete and partial IS elements with >85% identity at the DNA sequence level were considered to be the same as previously described IS elements. For the remaining elements, the highest database match at the aa sequence level was considered the closest relative. Only complete IS elements were given new IS number designations.

An intact copy of IS100 is located downstream of yopH in pCD1 (Fig. 2). There are numerous copies of IS100 throughout the genome of *Y. pestis* KIM strains ; the IS100 element (bp 12,609 to 14,562) in pCD1 (bp 12,609-14,562 of SEQ ID NO:1) is 100% identical in size and nucleotide sequence to a copy of IS100 present on the pesticin plasmid of *Y. pestis* strain EV76-6. A five base pair direct repeat flanks the IS100 which appears to have inserted within the relic of another insertion element. Five and seven base pair duplications have been found flanking other IS100 elements in *Y. pestis*.

IS1616 is a new 1,254 bp insertion element located at bp 50,753 to 51,987 of the entire assembled sequence, between ylpA and the sopABC partitioning region. The inverted repeats at the ends of IS1616 are 40 bp long and contain 9 mismatches. No direct repeats were detected flanking this element. While some elements do not generate a direct repeat upon transposition, the absence of direct repeats could be indicative of changes in the flanking DNA as a result of mutations that have occurred over time. There are

three open reading frames within IS1616, the first ORF (OrfA, bp 50,825 to 51,142) is predicted to encode a protein of 105 aa with a pI of 12.6. A second ORF of 186 aa (OrfB, bp 51,064 to 51,624) overlaps OrfA in the -1 frame. An additional 101 aa (orfC, bp 51,625 to 51,930), which may have originally been part of the second ORF, are encoded in the same frame just past the stop codon at bp 51,622 for OrfB.

IS1617 is a new 1,214 bp element, with inverted repeats of 39 and 40 bp containing 13 mismatches, located downstream of sycH. The five bases flanking each end of IS1617 are identical in 4 out of 5 positions. Like IS1616, this element belongs to the IS3 family and contains 2 overlapping ORFs with OrfB in the -1 frame relative to OrfA. OrfA could encode an 88 aa protein (bp 62,202 to 62,468, complement) while OrfB is open for 289 aa (bp 61,369 to 62,238, complement). A potential translational frameshift window of AAAAAAG is present in OrfA. IS1617 is more closely related to IS1222 from *Enterobacter agglomerans* and to ISD1 found in *Desulfovibrio vulgaris* than to IS1616. A remnant of IS1617 is present downstream of yopJ in pCD1 as well as in *Y. pseudotuberculosis* pIB1.

We found no evidence for the existence of yopL and, in *Y. pestis*, ylpA and yadA are pseudogenes. Although regulatory and secretory components of the LCR constitute a contiguous LCR cluster, elements suggesting this region is a pathogenicity island were not identified. Effector Yops are scattered throughout the plasmid and have widely varying GC contents, indicative of multiple gene acquisition events. This observation coupled with the presence of IS remnants from only distantly related microorganisms suggest a very complex history of DNA acquisition, insertions, deletions, and rearrangements was required for assembly of pCD1.

We failed to find genes with similarities to putative virulence factors that are not potential members of the LCR. However, we did identify eight ORFs of unknown function (Orfs 5,

59-61, 73, 74, 84, and 85). Orfs 7, 42-44, and 75 as well as YopT and its chaperone SycT are potential new members of the LCR virulence system. Sequence analysis of Orf7 suggests that it could be a chaperone for YopJ. Further investigation of these Orfs will allow assignment of their functions as LCR members or non-LCR virulence determinants.

We corrected the sequence of yopM, showing that it has two additional LRR repeats that are absent in *Y. enterocolitica*. While most LCR-related Y. pestis gene products showed 98% identity to their analogous *Y. enterocolitica* gene products, YopJ, YscG, YscE were ~94% identical to *Y. enterocolitica* products. It will be necessary to determine whether any of the differences in YopM, YopJ, YscG, YscE and the lack of a functional YlpA gene product are involved in differing levels of virulence among the pathogenic yersiniae.

An analysis was also done of the ORFs present in pPCP1. This analysis is presented in Table 4 below.

TABLE 4

Gene ID	Coords.	Genpept Gi#match	Description of Match
Y0002	971>1165	gi|455143	RNA I inhibition modulator protein (rom)
Y0003	1532>1903	gi|144312	ORF [Plasmid ColE1]
Y0004	2389>2826	gi|1200166|gnl|PID|e223344	pesticin immunity protein [Yersinia pestis]
Y0005	2861<3934	gi|984824	pesticin [Yersinia pestis]
Y0006	4052>4468		unknown
Y0007	4711>5649	gi|155525	plasminogen activator [Yersinia pestis]
Y0008	5836<6135	gi|1806206|gnl|PID|e293663	unknown [Mycobacterium tuberculosis]
Y0009	6135<6482		unknown
Y0010	7312<7686		unknown
Y0011	7743>8765	gi|1655837	ORFA; putative transposase

[Yersinia pestis]

Y0001 8762>9544 gi|1655838 ORFB; putative transposase [Yersinia pestis]

Thus the genes Y004, Y005 and Y007 are of particular interest as targets for use in treatment strategies due to their relationship with pathogenicity.